

Face-scanning behavior to silently-talking faces in 12-month-old infants: The impact of pre-exposed auditory speech

Claudia Kubicek¹, Anne Hillairet de Boisferon², Eve Dupierri², Hélène Lœvenbruck^{2,3}, Judit Gervain^{4,5}, and Gudrun Schwarzer¹

1-Department of Developmental Psychology, Justus-Liebig-University Giessen, Germany

2-Laboratoire de Psychologie et NeuroCognition, Université Pierre Mendès, Grenoble, France

3-Speech and Cognition Department - GIPSA-lab, UMR CNRS 5216, Grenoble, France

4- Université Paris Descartes, Sorbonne Paris Cité, Paris, France.

5-CNRS, Paris, France

Correspondence concerning this article should be addressed to Claudia Kubicek, Abteilung für Entwicklungspsychologie, Justus-Liebig-Universität Giessen, Otto-Behaghel-Str. 10 F1, 35394 Giessen, Germany; E-Mail: claudia.kubicek@psychol.uni-giessen.de

Running head: 12 month-olds' face-scanning behavior

Abstract

The present eye-tracking study aimed to investigate the impact of auditory speech information on 12-month-olds' gaze behavior to silently talking faces. We examined German infants' face-scanning behavior to side-by-side presentation of a bilingual speaker's face silently talking German utterances on one side and French on the other side, before and after auditory familiarization with one of the two languages. The results showed that 12-month-old infants showed no general visual preference for either of the visual speeches, neither before nor after auditory input. But, infants who heard native speech decreased their looking time to the mouth area and focused longer on the eyes compared to their scanning behavior without auditory language input, whereas infants who heard non-native speech increased their visual attention on the mouth region and focused less on the eyes. Thus, it can be assumed that 12-month-olds quickly identified their native language based on auditory speech and guided their visual attention more to the eye region than infants who have listened to non-native speech.

From birth on, infants encounter a socially rich environment, hearing and seeing talking faces on a daily basis, especially when communicating with their caregivers. Such early face-to-face communication should lead to a close binding between face and speech. It provides infants with redundant or complementary auditory-visual speech information, which could enhance their perception, understanding and learning of spoken language.

Although there is a substantial body of literature on infants' auditory processing of speech (e.g., Werker & Tees, 1999), far less researchers have focused on the visual information that infants extract from talking faces. Weikum and collaborators (2007) addressed the question whether visual speech cues alone are sufficient for infants' language discrimination. The authors used silent video clips of English-French bilingual speakers who were reciting different sentences in either French or English in a habituation paradigm. Monolingual 4- and 6-month-old infants were able to visually discriminate between the two languages, whereas infants at 8 months of age were not. The decline in the ability to visually discriminate their native language from an unfamiliar language observed in 8-month-olds is concordant with the perceptual narrowing view (Scott, Pascalis, & Nelson, 2007), that is, a tendency for infants to still develop perceptual abilities for native attributes, while declining for non-native attributes that infants have little experience with. This view has been described, for example, in studies of infants' phonetic development (Kuhl, Tsao, & Liu, 2003; Werker & Tees, 1984), face discrimination (Kelly et al., 2007; Pascalis, de Haan, & Nelson, 2002), and scanning of own- and other-race faces (Liu et al., 2011; Wheeler et al., 2011).

A recent study by Pons and collaborators (2009) investigated infants' intersensory speech perception. The authors compared 6- and 11-month-old English and Spanish infants' ability to orient toward silent videos of a Spanish woman pronouncing the syllables "ba" or "va" before and after auditory-only familiarization with one of the two syllables. The /ba/ vs. /va/ phonological contrast is known to be perceived by adult English speakers but not by

Spanish ones. As expected, no visual preference for either of the silent videos was observed prior to auditory familiarization for any of the four groups of children. After the auditory-only familiarization with either the /ba/ or the /va/ syllable, 6-month-old English and Spanish infants looked longer at the matching visible syllables, but at 11 months of age only the English infants did so. The fact that 6-month-olds looked at the corresponding visual stimulus more than at the non-corresponding stimulus suggests that these infants have perceived the relationship between the auditory and the visual stimulus, and thus that they performed intersensory matching. The fact that older infants did not match audible and visible attributes of non-native speech is interpreted by Pons and colleagues as suggesting that infants' sensitivity to intersensory speech narrows during the first year of life and that perceptual narrowing is a domain general developmental process.

By tracking eye movements, it was demonstrated that the way in which infants process audible talking faces in their native language changes over the course of the first year: Lewkowicz and Hansen-Tift (2012) investigated infants' visual attention to facial regions in 4- to 12-month-old English-learners while they watched and listened to a female speaking their native (English) or a non-native (Spanish) language. The authors were particularly interested in potential differences regarding the face-scanning behavior at the end of the first year. Independently of language familiarity, 4-month-olds looked longer at the eye region, 6-month-olds gazed equally long at the eyes and the mouth, and 8-month-old infants looked longer at the mouth area. Interestingly, 12-month-olds gazed equally long at the eyes and the mouth when native speech was spoken, whereas they maintained their looking toward the mouth area for the non-native language. The authors concluded that perceptual narrowing and growing native language expertise explained this behavior. Whereas the processing of non-native speech required infants to access complementary audio-visual cues from the mouth region, the processing of native speech was less depending on such cues.

To sum up, on the one hand, monolingual 8-month-old infants seem to no longer be able to visually discriminate their native language from an unfamiliar language, an ability infants have at 4 and 6 months of age (Weikum et al., 2007). On the other hand, when auditory speech information is presented, 12-month-old infants show a differential scanning pattern depending on whether the faces talked in their native or non-native language (Lewkowicz & Hansen-Tift, 2012), and 11-month-olds seem to remain sensitive to visual cues to native speech contrasts after audible-only speech input (Pons et al., 2009). These findings lead to the general question of the extent to which auditory speech has an impact on the visual processing of silently talking faces at the end of the first year.

To address this question, we used a variant of the intersensory matching procedure (Pons et al., 2009) and compared 12-month-old German infants' gaze behavior to facial regions of faces silently talking German (native) and French (non-native) fluent speech before (baseline trials) and after (post-test trials) auditory-only familiarization with one of the two languages. In accordance with the study of Weikum and collaborators (2007) who did not observe any visual language discrimination abilities in 8-month-olds, 12-month-old infants also should not show a visual preference for either of the languages. Moreover, we predicted no difference in scanning patterns between the German or the French speaking face in the baseline trials. By contrast, in the post-test trials, corresponding with the results of Lewkowicz and Hansen-Tift (2012), we expected infants' looking times to the *mouth* region to increase when French auditory language is presented, in comparison to baseline, and infants' looking times to the *eyes* to increase when German auditory language is presented.

Method

Participants

Forty monolingual German-learning 12-month-old infants ($M = 377$ days, $SD = 8$ days; age range: 365 days – 394 days, 21 males) were tested. All infants were full-term with

no visual or auditory deficits, as reported by parents. The data from seven additional infants were discarded from the final sample due to equipment failure (two), failure to complete the calibration procedure (four) or extreme fussiness (one).

Stimuli

Visual stimuli were silent video clips of four female bilingual German-French speakers. Recording took place in Germany (two speakers) and in France (two speakers). The speakers were recorded against a blue background looking directly into the camera while they were reciting a German and a French version of a children's short story in a child-directed manner. Each of the 30-second video clips showed a full-face image of the speaker and measured 20.6 cm x 18 cm when displayed side-by-side on the monitor, separated by an 11-cm gap. Both videos, French and German, were edited to make sure that they started on a closed mouth and the first mouth opening was synchronized.

Auditory stimuli consisted of the same story and were presented at conversational sound pressure level ($65\text{dB} \pm 5\text{dB}$).

Apparatus and Procedure

Caregivers were informed of the purpose of the study and gave written consent for their infant to participate. Each infant was tested individually in a baby lab, the caregiver sitting on a chair with the infant on her lap. To prevent parents from influencing the looking behavior of their infants, they were told to keep their eyes closed and to refrain from talking for the duration of the experiment. The infants were seated on the caregiver's lap at a distance of 60 cm in front of a 22-inch monitor (resolution: 1280 x 1024 pixels). Stimuli were presented by using E-Prime 2.0 software (Psychology Software Tools, Sharpsburg, PA).

Eye-tracking data were captured by a Tobii X120 eye tracker with a sampling rate of 60 Hz. Infants completed an infant-adapted 5-point calibration sequence, which was run using

Tobii Studio software (Tobii Technology, Sweden). Calibration was checked for accuracy and repeated up to three times if necessary.

After showing an attention getter, we used a procedure adapted from Pons et al. (2009), which is a variant of the intersensory matching procedure. Importantly, in this procedure the sound is not presented at the same time as the visual stimuli to ensure that audiovisual synchrony is not mediating intersensory matching.

Insert Figure 1 about here

There were six 30-second trials (see Figure 1): in the first and second trials (baseline condition), infants were presented with two side-by-side silent video clips, displaying one bilingual speaker uttering the same story in French on one side and in German on the other side. The left-right position of French and German videos was counterbalanced across infants in the first trial and reversed in the second one. In the third trial (auditory familiarization trial), infants were presented with a voice speaking either French or German while they were watching an attention getter. Infants were randomly assigned to one of two auditory groups (German or French) and the specific speaker the infants listened to was different from the speaker they had seen in the baseline trials. In the 4th trial (post-test trial), we presented the two initial silent videos again. The 5th and 6th trials were a repetition of the auditory familiarization and post-test trial but the left-right presentation of the silent videos was reversed in the 6th trial. The voices and silent videos of the four female bilingual German-French speakers were counterbalanced across infants.

Data Analysis

Data were analyzed for the total duration of fixations within an area of interest. Fixations were defined as having a minimum radius of 35 pixels and a minimum duration of 100 ms. Three Areas of Interest (AOIs) were created: whole face, mouth, and eyes. The AOI of the whole face was defined by outlining the face with the hair excluded. A small buffer zone of approximately 1 cm allowed for head movements during the recording (see Liu et al.,

2011; Wheeler et al., 2011). The mouth AOI and the eyes AOI were created in the same way as in the study of Lewkowicz and Hansen-Tift (2012) by defining rectangular AOIs that covered both eyes and the mouth, respectively, for each stimulus face (see Figure 2).

Insert Figure 2 about here

For the three AOIs (face, mouth and eyes) of each French or German talking face, gaze data were aggregated across the baseline and post-test trials, respectively. Thus, data analysis was conducted with the AOIs (face, mouth and eyes) of the bilingual person talking French (resp. German) during the baseline trials (1st and 2nd trial), and the AOIs of the bilingual person talking French (resp. German) during the post-test trials (4th and 6th trial).

Results

Preliminary analyses revealed no significant effects of infants' gender or of the bilingual speakers' identity on infants' gaze behavior to AOIs. Thus, the data for these two factors were collapsed for subsequent analyses.

Face AOI.

As dependent variable we computed a preferential looking score by dividing the fixation time on one face (e.g. face AOI of the face talking German) by the amount of total looking time (sum of looking times at the face AOI of the German and French faces) across the baseline and post-test trials, respectively. To determine whether the infants showed a preference for one of the visual speeches during the baseline condition and whether such a preference could be affected by the auditory speech the infants were familiarized with, we submitted the preferential looking scores to an ANOVA with “visual speech” (German, French) and “test” (baseline, post-test) as within-subjects factors, and “auditory group” (French, German) as a between-subjects factor.

This analysis did not show any significant effect (all $ps > .6$), indicating that infants did not show a preference for either of the faces during baseline ($M_{German} = 50\%$, $SD_{German} = 11\%$; $M_{French} = 50\%$, $SD_{French} = 11\%$) and that this pattern was not influenced by auditory familiarization, neither after infants were exposed to German auditory speech ($M_{German} = 48\%$, $SD_{German} = 10\%$; $M_{French} = 52\%$, $SD_{French} = 11\%$) nor after the infants heard French ($M_{German} = 50\%$, $SD_{German} = 11\%$; $M_{French} = 50\%$, $SD_{French} = 11\%$).

Proportional fixation on mouth and eyes AOIs

The fixation data were converted into proportional fixation times within the mouth and eyes AOI relative to the total on-face fixation times. To analyze whether the infants showed different looking patterns within the faces after auditory familiarization, we computed a four-way ANOVA with “AOI” (mouth, eyes), “visual speech” (German, French) and “test” (baseline, post-test) as within-subjects factors and “auditory group” (French, German) as a between-subjects factor.

This analysis failed to show any effect of visual speech, $F(1, 38) = .04$, $p = .83$, $\eta^2 = .00$, nor any interaction with this factor. This seems to indicate that infants scanned the German and French speaking faces in a similar way, even after auditory familiarization. However, the ANOVA revealed a significant main effect of AOI, $F(1, 38) = 126.23$, $p < .001$, $\eta^2 = .77$. Infants looked longer at the mouth AOI ($M_{mouth} = 48\%$, $SD_{mouth} = 20\%$) than at the eyes AOI ($M_{eyes} = 14\%$, $SD_{eyes} = 12\%$). We also found a significant AOI x auditory group interaction, $F(1, 38) = 9.78$, $p < .01$, $\eta^2 = .21$ and a significant AOI x test x auditory group interaction, $F(1, 38) = 24.87$, $p < .001$, $\eta^2 = .40$, indicating that during the post-test trials the looking patterns toward the mouth and the eye region of both the German and French speaking faces were influenced by the auditory input. However, the looking time was affected similarly for German and French faces, as there was no difference in looking times *between* both faces (see above). As illustrated in Figure 3, this latter interaction is explained by the fact that the

auditory familiarization modulated looking behavior in an opposite manner for French and German.

Insert Figure 3 about here

Whereas infants auditorily familiarized with French looked longer at the mouth region of both faces during the post-test trials ($M = 57\%$, $SD = 15\%$) than during baseline ($M = 48\%$, $SD = 11\%$; $t[19] = -2.18$, $p < .05$), infants who had heard German looked shorter at the mouth region of both faces during the post-test trials ($M = 35\%$, $SD = 14\%$) than during baseline ($M = 48\%$, $SD = 17\%$; $t[19] = 5.09$, $p < .001$). Conversely, whereas infants auditorily familiarized with French looked shorter at the eyes region of both faces during the post-test trials ($M = 9\%$, $SD = 5\%$) than during baseline ($M = 14\%$, $SD = 8\%$; $t[19] = 2.23$, $p < .05$), infants who heard German looked longer at the eyes region of both faces during the post-test trials ($M = 23\%$, $SD = 9\%$) than during baseline ($M = 14\%$, $SD = 10\%$; $t[19] = -3.42$, $p < .01$).

Discussion

The main goal of the present study was to investigate whether 12-month-olds' visual exploration of faces talking in their native language (German) or in a non-native (French) language was influenced by auditory speech. The results showed that listening to auditory speech did not lead to different visual preferences for the German and French talking faces. Auditory speech rather affected infants' visual scanning of the faces but independently of whether they were silently talking French or German. After auditory exposure to their native language, infants looked less at the mouth of *both* faces, whereas looking times at the eyes increased. Conversely, after auditory exposure to their non-native language, infants looked more at the mouth of *both* faces, whereas looking times at the eyes decreased.

These results are consistent with the study of Lewkowicz and Hansen-Tift (2012) who showed that 12-month-old infants looked longer at the mouth when they heard and saw a

person talking in a foreign language (Spanish), in contrast to infants who attended equally long on the mouth and the eyes when the talker spoke their native language (English). However, Lewkowicz and Hansen-Tift (2012) presented the infants with visual and auditory speech simultaneously whereas our design allowed us to isolate the impact of separately presented auditory speech on infants' looking patterns to silently talking faces.

According to Lewkowicz & Hansen-Tift (2012), our results can be considered in the context of multisensory perceptual narrowing. As infants at the end of the first year have specialized speech perception skills of their native language (Werker & Tees, 1999), it can be assumed that the processing of their native speech does not necessarily rely on language specific visual inputs like the movements of the mouth. On the contrary, the processing of unfamiliar speech relies more on the processing on the mouth. In line with this interpretation our results showed that after hearing their native language, infants focused less on the mouth and increased looking times toward the eyes of the silently talking faces, probably to gain access to socially relevant cues (Langton, Watt, & Bruce, 2000). In contrast, infants who were presented with French auditory speech (their non-native language) focused less on the eyes and showed increased visual attention to the mouth, possibly to profit from complementary audio-visual information that is available in this area.

With regard to infants' visual processing of silently talking face without previous listening to auditory speech (baseline trials) our results have demonstrated that the infants did not show any preference for one of the two visual speeches, suggesting that they were not able to discriminate their native language from a non-native language on the basis of visual cues alone. These results confirm and extend the findings of Weikum and collaborators (2007) by showing similar results among infants older than 8 months. Unexpectedly, our findings showed that infants focused longer on the mouth than on the eyes on both the German and the French speaking faces in the baseline (see Figure 3). Thus, without the opportunity to listen to

speech, infants at 12 months of age looked mainly at the mouth compared to the eyes of silently talking faces. Our results are in line with studies indicating that older infants focus longer on the mouth of in-sound presented talking faces compared to smiling faces (Tenenbaum, Shah, Sobel, Malle, & Morgan, 2012), or when they are exposed to silently talking faces of young children (Frank, Vul, & Saxe, 2011) or of their mother (Hunnius & Geuze, 2004). It is reasonable to assume that in our study, infants' attention was mainly attracted by the movement of the mouth when no auditory speech was presented before.

Importantly, the auditory input did not enable the 12-month-old infants to discriminate the visually presented French and German languages neither in terms of a general visual preference nor in terms of differential scanning patterns between the faces. Thus, even for their native language, infants did not perform cross-modal matching of heard and seen speech. This result seems to be in contrast to Pons and collaborators' result (2009) indicating that 11-month-old English infants successfully discriminated between the visually presented syllables "ba" and "va" after they had heard one of these syllables acoustically. Thus, it seems that, at the syllable level, discrimination of visual speech or cross-modal matching of visual-auditory information can be enabled by prior auditory speech in infants at the end of the first year. At the utterance level, however, auditory speech does not enable visual language discrimination. It can be assumed that cross-modal processing of fluent speech, especially when amodal cues (e.g. synchrony) are eliminated due to sequential presentation of auditory and visual speech, is too difficult for infants at this age. Another interpretation could be that infants at the end of the first year do not notice the differences between the visual cues carried by two different languages at the utterance level.

A limitation of the present study is that only 12-month-olds were tested and future research will have to be conducted using our design to examine younger infants. This will allow us to determine the extent to which higher interest in visual speech combined with less

auditory language understanding, that can be assumed in younger infants, affect their visual processing of silently talking faces.

References

- Frank, M. C., Vul, E., & Saxe, R. (2011). Measuring the development of social attention using free-viewing. *Infancy*, 16, 1-21.
- Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, 6(2), 231-255.
- Kelly, D.J., Quinn, P.C., Slater, A.M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, 18(12), 1084-1089.
- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences USA*, 100, 9096-9101.
- Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4(2), 50-59.
- Lewkowicz, D., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences USA*. doi: 10.1073/pnas.1114783109
- Liu, S., Quinn, P. C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and difference in the processing of same- and other-race faces as revealed by eye tracking in 4- to 9-month-olds. *Journal of Experimental Child Psychology*, 108, 180-189.
- Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is Face Processing Species-Specific During the First Year of Life? *Science*, 296, 1321-1323.
- Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences USA*, 106, 10598-10602.

- Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, 16(4), 197-201.
- Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2012). Increased focus on the mouth among infants in the first year of life: A longitudinal eye-tracking study. *Infancy*, 17, 1-20.
- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, 316, 1159.
- Werker J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.
- Werker J. F., & Tees, R. C. (1999). Influences in infants speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509-535.
- Wheeler, A., Anzures, G., Quinn, P. C., Pascalis, O., Omrin, D. S., & Lee, K. (2011). Caucasian infants scan own- and other-race faces differently. *PLoS ONE*, 6, e18621. doi:10.1371/journal.pone.0018621

Funding

This study was supported through a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft) for Gudrun Schwarzer(SCHW 665/11-1) and ANR--10-FRAL-017 for HL & JG.

Figure 1. Schematic representation of the procedure used in the current study. Only the German auditory condition is shown.

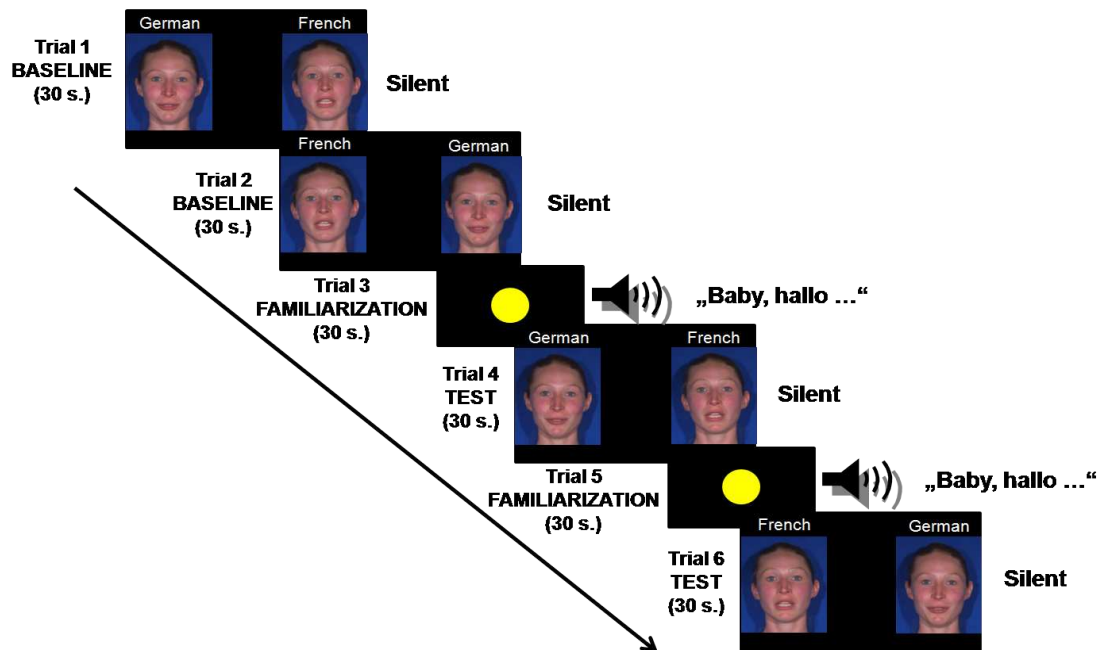


Figure 2. Example of eyes and mouth Area of Interest (AOI) plots.

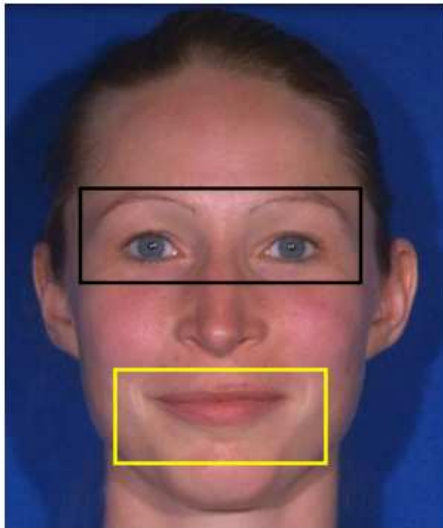


Figure 3. Mean and standard errors of proportional looking times (%) to mouth and eyes AOI within baseline and post-test trials of both auditory groups.

